EXHIBIT D

3GPP TSG-RAN WG2 Meeting #57 **R2-070476**
12 - 16 February 2007
St Louis, Missouri, USA

Agenda item: **11.8**

Source: **Nokia**

Title: **Uplink Scheduling for VoIP**

Document for: **Discussion and Decision**

1 Introduction

Some scheduling alternatives for UL VoIP are described and discussed.

2 Scheduling schemes for UL VoIP

2.1 Dynamic scheduling of UL VoIP

Since LTE is a packet radio system where normally each packet is scheduled by L1/L2 control signalling, it is natural to apply this dynamic scheduling also for VoIP packets as much as possible. Dynamic scheduling of each VoIP packet/transmission is naturally most flexible from the scheduling and UL resource usage point of view but also requires most signalling. The fully dynamic scheduling means that the UE sends a resource request in UL for every VoIP packet (UE could tell even the VoIP packet size), Node B allocates UL resource for every VoIP packet separately and for every retransmission separately. With dynamic scheduling VoIP users can benefit from time/frequency selective scheduling and unused resources due to silent periods as well as due to early termination of HARQ can be easily reallocated to other VoIP users, especially if adaptive HARQ is used (retransmissions can be allocated on different RUs).

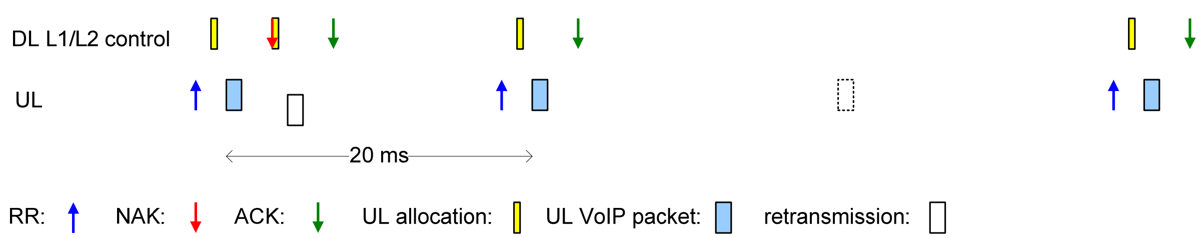

Figure 1: Dynamic scheduling of UL VoIP

The amount of downlink signalling can be reduced by not scheduling the retransmissions but instead using the same (or other predefined) resource for retransmissions, however, ACK/NAK signalling is needed anyway. If only the first transmissions are scheduled, then HARQ has to be synchronous and non-adaptive. The disadvantage is that the scheduling flexibility of the retransmissions is lost and that results into 'holes' in the time/frequency space when some retransmission resources are unused and others cannot be reallocated.

The amount of uplink signalling can be reduced by sending the resource request only at the beginning of the talk spurt. The Node B then allocates the uplink resource dynamically about every 20 ms. Similarly, during the silent periods (which the UE indicates explicitly or implicitly), the allocation is given every 160 ms.

Pros:

- Flexible scheduling of VoIP and other users, especially if all the transmissions scheduled dynamically

- Scheduling freedom, no predetermined split in resources between VoIP users and other users
- Frequency and time selective scheduling possible (may require more channel sounding pilot signalling)
- Fast and slow link adaptation possible
- Asynchronous HARQ is possible and retransmission resources can be allocated freely to VoIP or non-VoIP users
- uplink resource request only at the beginning of the talk spurt saves uplink resources

Cons:

- Amount of L1/L2 signalling required
- With synchronous non-adaptive HARQ retransmission resources difficult to allocate to other users
- Allocation of the TB size has to be done for the maximum VoIP packet size, i.e., there is padding (with SID frame unto 30 bytes)

2.2 Persistent allocation of VoIP

Some sort of persistent allocation for VoIP users has been proposed by many companies. The main driving force is to reduce the amount of L1/L2 control signalling or even get rid of it completely and signal the allocation instead by using RRC signalling.

A fully persistent allocation would mean a CS like allocation for VoIP. RRC signalling would be used to allocate a time/frequency resource (localized or distributed) as well as a fixed modulation scheme to a VoIP user. The allocation should also include resources required for HARQ retransmissions. The allocation could even be so persistent that HARQ ACK/NAKs are not sent but instead each packet is sent a fixed number of times (as proposed in [1]), thus resulting in a fixed FEC scheme instead of an adaptive HARQ scheme.

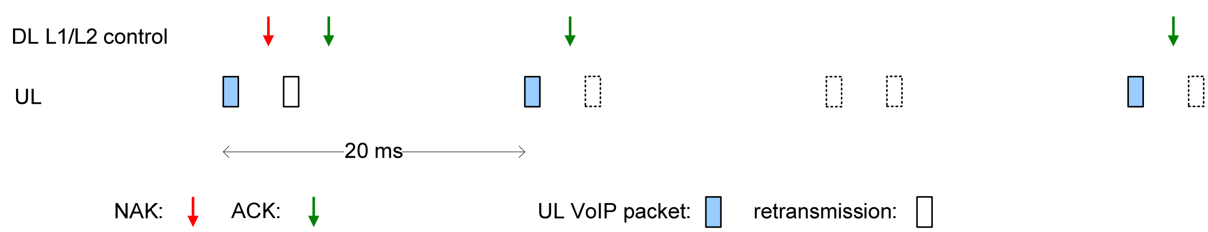

Figure 2: Persistent allocation of UL VoIP

Pros:

- Reduced signalling overhead (no L1/L2 allocation signalling in DL, only ACK/NAKs)
- simple

Cons:

- Wasted resources
    - Unused HARQ retransmission resources not used for other users
    - Unused resources during voice silent periods not allocated to others
- A DTX detector needed at Node B, otherwise Node B tries to decode and combine the transmissions in pre-allocated resources, even when nothing is sent
- Persistent allocation of the TB size has to be done for the maximum VoIP packet size, i.e., there is padding (with SID frame unto 30 bytes padding)
- Max VoIP (and system) capacity limited since unused resources cannot be allocated to other users

- Requires non-adaptive HARQ

2.3 Semi-persistent scheduling of UL VoIP

Since VoIP users are on average half of the time silent, a significant amount of VoIP capacity is wasted if the silence periods are not reallocated to other VoIP users. Therefore, even the persistent allocations should be such that silence periods can be reallocated to other VoIP users. This is possible if the persistent allocation is done separately for each talk spurt. When a talk spurt starts, the UE should send a resource request, then the radio resource is allocated to the UE and when the talk spurt ends, the resource is released (explicitly with release signalling or implicitly by noticing that no more data is coming). Thus the released resource can be allocated to some other VoIP user.

SID frames could be allocated dynamically (thus requiring RR for each SID) or another persistent allocation could be given for them.

At the beginning of the talk spurt (UE RLC/MAC notices that there is full rate speech packet coming) UE should send an uplink resource request (RR) to the Node B. The RR can be sent either on synchronous RACH, on CQI channel or on a dedicated RR channel.

At the beginning of a talk spurt the UE is allocated a semi-persistent time/frequency resource where the UE can send the initial transmissions without receiving UL allocation via the L1/L2 control channel. The semi-persistent allocation is sent either on L1/L2 control channel, in a MAC control PDU or as an RRC message. All the retransmissions are allocated dynamically using the L1/L2 control channel (NAK is always followed by an UL allocation). The UE monitors the L1/L2 control channel in all or in preconfigured TTIs (DRX). If no valid UL allocation is given to the UE, the UE is allowed to send an initial data transmission using the pre-assigned resource (using pre-assigned transport format). Since the retransmissions are always scheduled, this scheme allows using adaptive HARQ: the retransmissions can be freely allocated on any free resources, e.g., on those remaining unused by silent users.

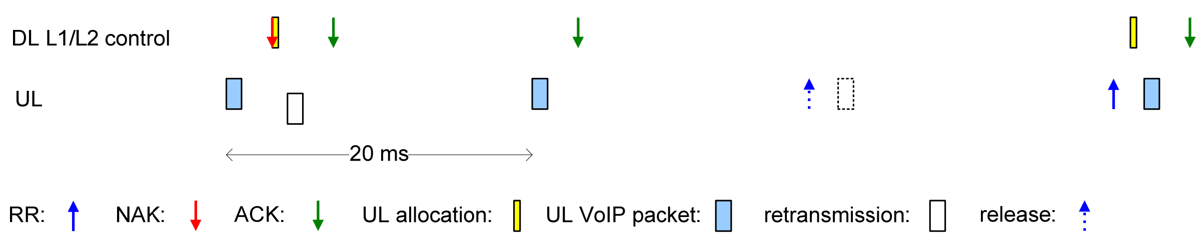

Figure 3: Semi-persistent allocation of UL VoIP

Semi-persistent allocation of UL VoIP is illustrated in Figure 3.

The persistent allocation methods typically assume synchronous non-adaptive HARQ scheme and do not allow reallocation of unused HARQ retransmission resources to other VoIP users. By increasing L1/L2 signalling somewhat, also the HARQ resources can be reallocated. This allows using adaptive HARQ for VoIP, too.

Pros:

- Slow talk spurt based link adaptation possible, even some frequency selective scheduling possible for slowly moving users
- Resources during silent periods released and can be reallocated to other (persistent) VoIP users, even a frequency hopping allocation can easily been reallocated
- Reduced DL L1/L2 control signalling only at the beginning of the talk spurt, SID frames may be scheduled dynamically or persistently
- Resource request (RR) signalling only at the beginning of the talk spurt (and for SID frames?)
- Allows using adaptive HARQ
- If the BLER target of the first transmission is in the order of 10-20%, only 10-20% of the transmissions require L1/L2 control signalling.
- Allows reallocating the unused HARQ resources to other VoIP users

Cons:

- Some capacity loss due to semi-persistent allocation of the initial transmission
- Persistent allocation of the TB size has to be done for the maximum VoIP packet size, i.e., some padding is needed (with SID frame unto 30 bytes padding unless RR can tell that a SID frame is in the buffer)

2.4 Discussion

Due to the benefits of dynamic scheduling we propose to use it as much as possible for VoIP. The scheduler can always start with dynamic scheduling. When VoIP signalling load increases, part of the VoIP users can be given a semi-persistent allocation for the initial transmissions. The retransmissions would still be scheduled for all users.

3 Simulation results

In this section a summary of the initial simulation results for the described schemes are presented. Simulation assumptions and further details of the results are presented in Annex A.

Table 1 summarises the VoIP capacity results for 7.95 kbps AMR codec in Case 4 and Case 1 [2] with 1.25 MHz bandwidth. The VoIP capacity for 12.2 kbps AMR codec with 1.25 MHz bandwidth is summarised in Table 2. The dynamic allocation is based on a DL control channel limitations of 3 users scheduled per TTI. Persistent allocation for both initial and retransmissions gives only about half the capacity provided by dynamic or semi-persistent allocation. It should be noted that in the simulated persistent allocation scheme resources are released during silent periods.

Table 1: VoIP capacity (users/sector), 7.95 kbps AMR codec, 1.25 MHz

	Case 4, without SID	**Case 4, with SID**	**Case 1, with SID**
Dynamic allocation	93	86	78
Semi-persistent with sync adaptive HARQ	89	82	77
Persistent sync non-adaptive HARQ (1 retrans)	44	-	-

Table 2: VoIP capacity (users/sector), 12.2 kbps AMR codec, 1.25 MHz

	Case 4, with SID	**Case 1, with SID**
Dynamic allocation	73	65
Semi-persistent with sync adaptive HARQ	71	63

The VoIP capacity for 12.2 kbps AMR codec on 5 MHz carrier bandwidth is shown in Table 3. The capacity of dynamic allocation depends on the number of allowed control channels per TTI. Around 6-7 UEs can be scheduled per TTI if the DL control channel uses n=3 OFDM symbols. To be able to schedule 10 UEs per TTI n should be increased to 4 or 5.

Table 3 VoIP capacity (users/sector), 12.2 kbps AMR codec, Case 1, 5 MHz

	Case 1, with SID
Dynamic allocation (7 UEs/TTI)	197
Dynamic allocation (10 UEs/TTI)	240
Semi-persistent with synchronous adaptive HARQ (10 UEs/TTI)	236

In all cases, rhe baseline provides the largest capacity, but without the drawback related to a semi-persistent kind of allocation [3].

4 Conclusions

In this contribution, different scheduling approaches for UL VoIP were described, discussed and compared through system level simulations. As shown for the DL [3], the baseline together with a proper setting of the number of control symbols provides the largest VoIP capacity. The relation between the number of control symbols, number of scheduled users and maximum VoIP capacity should therefore be discussed together with RAN1.

References

[1] R2-062164, *Uplink Resource Allocation Scheme*, NTT DoCoMo

[2] 3GPP TR 25.814

[3] R2-070475, *Downlink Scheduling for VoIP*, Nokia

Annex A

A1 Simulation Setup

The deployment scenarios are listed in Table A1: Deployment Scenarios.

Table A1: Deployment Scenarios

Scenario	CF (GHz)	ISD (m)	BW (MHz)	PLoss (dB)	Speed (km/h)	Propagation Model (R in Km)
Case 1	2	500	1.25/5	20	3	L = 128.1 + 37.6 Log10R
Case 4	0.9	1000	1.25	10	3	L = 120.9 + 37.6 Log10R

The overall system configuration is shown in Table A2: System simulation parameters.

Table A2: System simulation parameters

Parameter	Configuration
Layout	Hexagonal grid, 19 cell sites, 3 sectors per site
Antenna pattern	70 deg (-3 dB) with 20 dB front-to-back ratio
Standard deviation of slow fading	8 dB
Shadowing correlation between cells / sectors	0.5 / 1.0
eNodeB/UE antenna gain	14 dBi / 0 dBi
eNodeB receiver	2 antennas
Thermal noise density	-174 dBm/Hz
Frequency re-use	1
Channel model	6-ray Typical Urban
Traffic model	AMR 7.95 Kbps and 12.2 Kbps (50%-activity 2-state markov model with 3 second average talk-spurt duration); ROHC and other overhead added, then 28 bytes per AMR 7.95 Kbps VoIP speech packet 40 bytes per AMR 12.2 Kbps VoIP speech packet 15 bytes per SID packet
Max UE Tx Power	21 dBm
Channel update	per sub-frame (0.5 ms)
TTI length	1 ms
Control overhead per TTI	11 long blocks per TTI for data (21% overhead)
HARQ	Max. num of Txs = 4 Num of HARQ processes = 3 Synchronous or asynchronous The same MCS for retransmissions Chase combining, Ack/Nack errors = 0%
Interference Control	Semi-static IC

Frequency band allocation and MCS	Assignment of 2 resource units (180 kHz per RU) for transmission of one packet, then {
	QPSK(ECR=2.1493) for AMR 7.95 Kbps
	QPSK(ECR=1.5824) for AMR 12.2 Kbps
	QPSK(ECR=3.3488) for SID packet };
	Or 3 RUs for transmission of one packet, then {
	QPSK(ECR= 3.2239) for AMR 7.95 Kbps
	QPSK(ECR= 2.0971) for AMR 12.2 Kbps
	QPSK (ECR= 5.5385) for SID packet }

L2S	AVI assuming practical FDE receiver and realistic channel estimation
Evaluation method	5% outage based on users having < 98% of its speech frames delivered successfully within 40 ms (PER<2%)

A2 Simulation Result

From subsection A2.1 to A2.7, we study capacity in 1.25MHz bandwidth (total 6 RUs for data and no RU reserved for control) and 2 RUs for transmission of each packet; in A2.8, 5MHz BW is studied and 4 RUs of total 25 RUs are reserved for UL control signaling, the restriction of number of scheduled users per TTI is considered in A2.8 as well.

A2.1 Comparison of different HARQ operating points

First, different HARQ operating points for 7.95Kbps AMR in Case 4 under semi-persistent scheduling and synchronous HARQ are compared. SID packets are modeled and dynamically scheduled. Different BLERs are got by adjusting the transmission power of users. The capacity results and distribution of retransmissions for different HARQ operating points are shown in Figure A1 and Table A3, respectively. It's clear that less retransmission implies better capacity. From Table 3, it can be seen that only about 10% packets need retransmission for the optimum case. So for semi-persistent scheduling, it means only a small part of packets (about 10% in this condition) need DL allocation for retransmissions thus the DL control signaling is greatly reduced. Take 80 users and ~10% BLER for example, only 0.21 average UL "grants" per TTI are needed for semi-persistent scheduling; however, under the same assumption, up to 2.26 average UL "grants" per TTI are needed for dynamic scheduling, which is about 10 times more than semi-persistent scheduling.

Please note that the results in following sub-sections are all based on the optimum HARQ operating points.

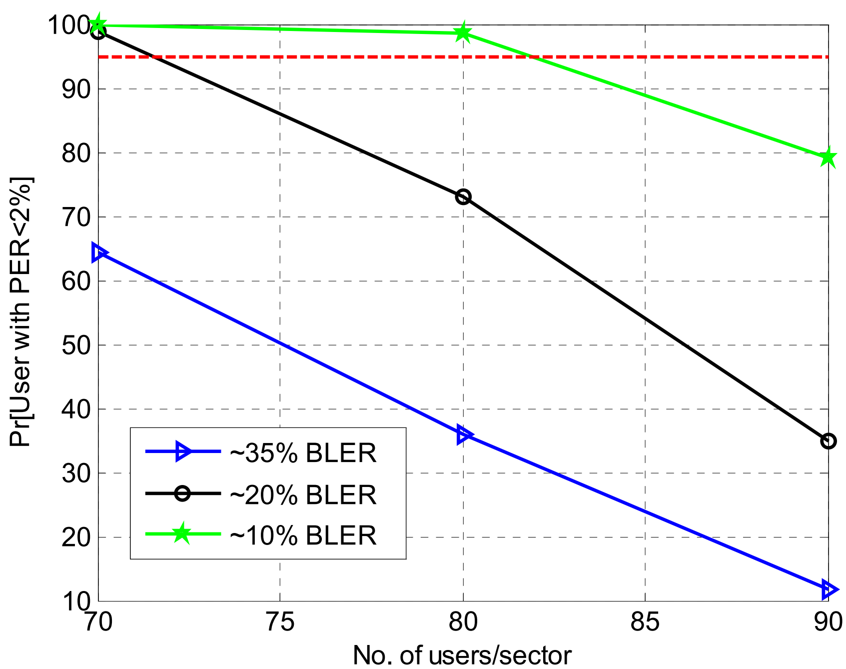

Figure A1 Capacity comparison of different HARQ operating points under semi-persistent scheduling

Table A3 Distribution of HARQ retransmissions for different BLER

BLER points	No. of UE/sector	Pr[succeed after 1st trans]	Pr[succeed after 2nd trans]	Pr[succeed after 3rd trans]	Pr[succeed after 4th trans]	Pr[failed packets]
~35%	70	0.6451	0.2687	0.0467	0.0079	0.0316
	80	0.6450	0.2211	0.0359	0.0057	0.0923
	90	0.6364	0.1408	0.0207	0.0029	0.1993
~20%	70	0.8172	0.1640	0.0159	0.0016	0.0012
	80	0.7861	0.1744	0.0183	0.0019	0.0193
	90	0.7791	0.1382	0.0132	0.0012	0.0683
~10%	70	0.9227	0.0728	0.0037	0.0004	0.0003
	80	0.9118	0.0816	0.0047	0.0006	0.0013
	90	0.8951	0.0841	0.0049	0.0005	0.0155

A2.2 UL VoIP capacity for persistent scheduling and 7.95Kbps AMR (without SID)

In this subsection, the capacity results for persistent scheduling are presented in Figure A2 UL VoIP capacity for persistent scheduling and 7.95Kbps AMR (without SID). SID frames are not considered (There are no packets in DTX periods). It can be seen that about 41 and 44 users per sector can be supported at 95% outage threshold in Case 1 and Case 4, respectively. The reason of such low capacity is that most resources which are reserved for retransmissions (due to the characteristic of persistent scheduling) are wasted due to early HARQ termination. This can be regarded as the baseline capacity for UL VoIP. The persitent scheduling is no longer evaluated in the following subsections due to its proved poor performace.

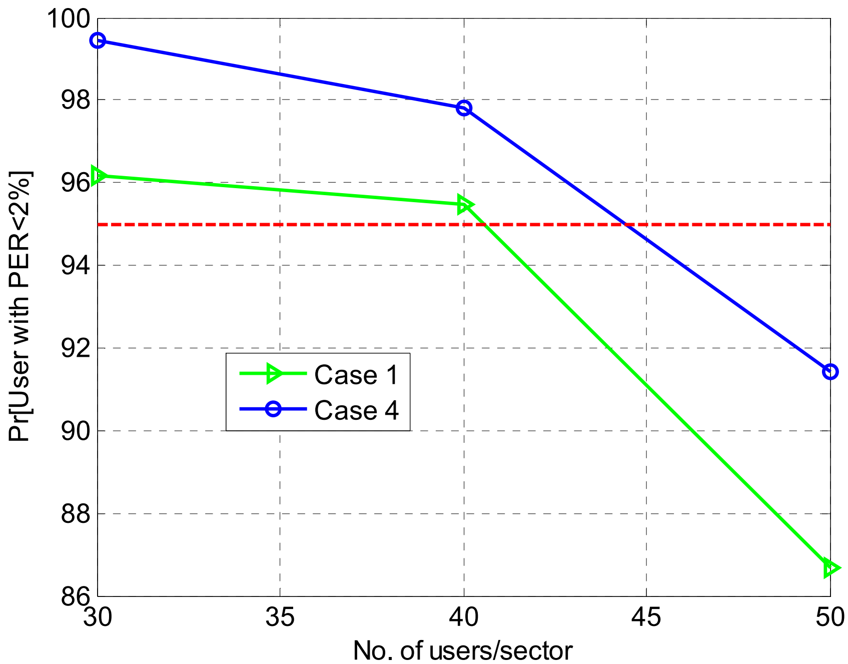

Figure A2 UL VoIP capacity for persistent scheduling and 7.95Kbps AMR (without SID)

A2.3 UL VoIP capacity for 7.95Kbps AMR in Case 4 (without SID)

The capacity results for semi-persistent scheduling and dynamic scheduling without SID modeling are presented in Figure A3 UL VoIP Capacity for 7.95Kbps AMR in Case 4 (without SID). It can be seen that about 90 users/sector can be supported, which is about 100% higher than pure persistemt scheduling due to more efficient resource use for retransmissions. Under light load (<=90 UE per sector), semi-persistent scheduling achieves almost the same performance as dynamic scheduling; under heavy load (100 UE per sector), performance of semi-persistent scheduling decrease sharply, this is because too many users are connected to eNodeB thus their initial transmissions occupy most resources and little resources are remained for retransmissions. The solution is to reserve some resources only for retransmissions but not for initial transmissions (control the number of admitted users per sector). From the figure, it shows that asynchronous HARQ achieves a bit greater capacity than synchronous HARQ due to more flexible retransmission in time at the cost of increased signalling.

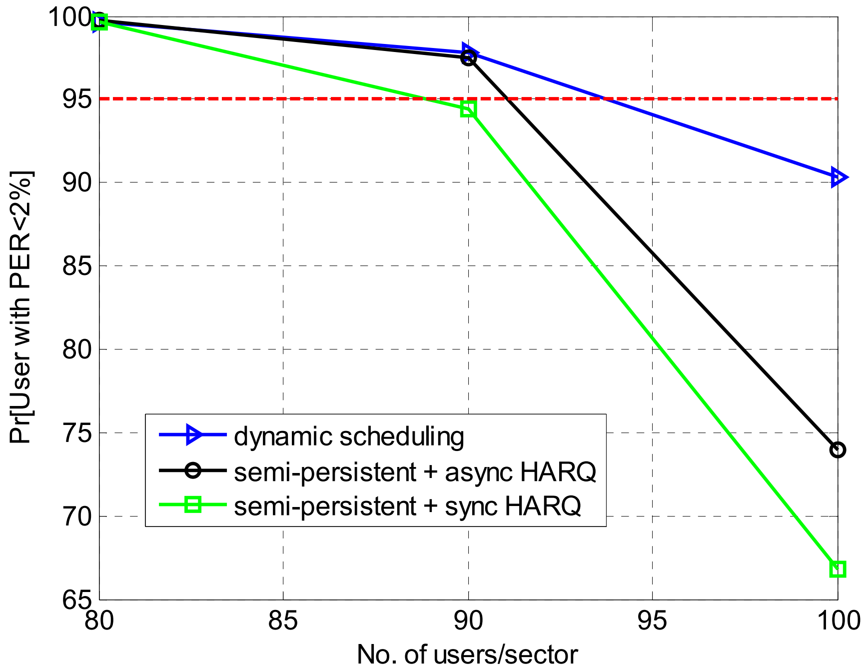

Figure A3 UL VoIP Capacity for 7.95Kbps AMR in Case 4 (without SID)

A2.4 UL VoIP capacity for 7.95Kbps AMR in Case 4 (with SID)

In this case, SID packets are modeled and dynamically scheduled. From Figure A4 UL VoIP Capacity for 7.95Kbps AMR in Case 4 (with SID), with SID packets modeled, 82-86 users per sector can be supported and there is a capacity reduction of nearly 10% compared to 89-94 users in Figure A2 where SID is not modeled.

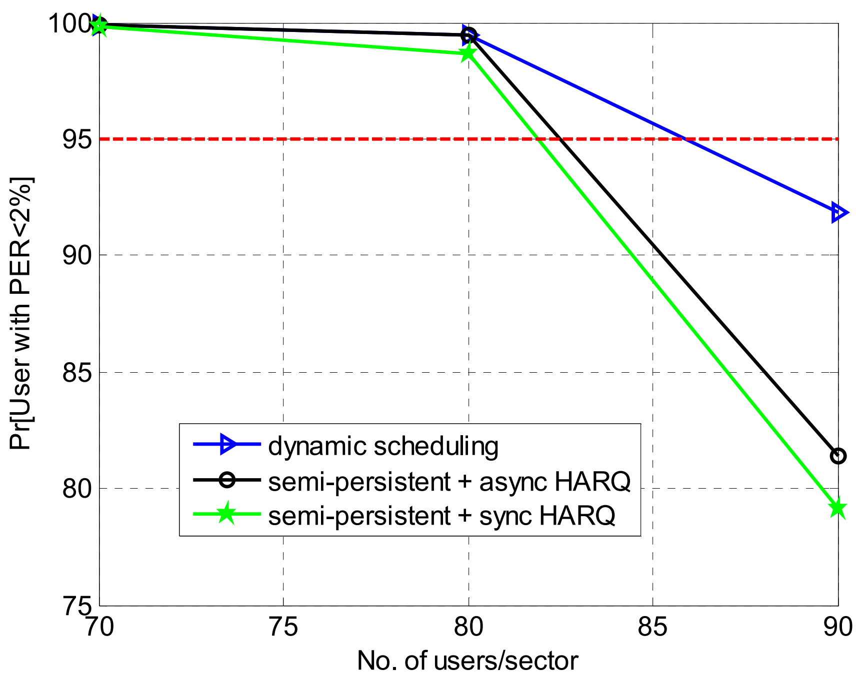

Figure A4 UL VoIP Capacity for 7.95Kbps AMR in Case 4 (with SID)

The delay distribution for different scheduling methods is shown in Figure A5 Delay CDF for 7.95Kbps AMR in Case 4 (with SID, 80 UE/sector). It assumes that eNodeB immediately knows it when UE has packets in buffer and when UE changes status (from active to DTX or from DTX to active). So the packets delay mainly includes two parts: the waiting delay until eNodeB finds free resource for UE due to system overloading and the transmission delay due to possible retransmissions. From the figure, dynamic scheduling gets better delay distribution than semi-persistent scheduling because retransmission has higher priority. Asynchronous HARQ gets better delay distribution than synchronous HARQ because retransmission is more flexible in time.

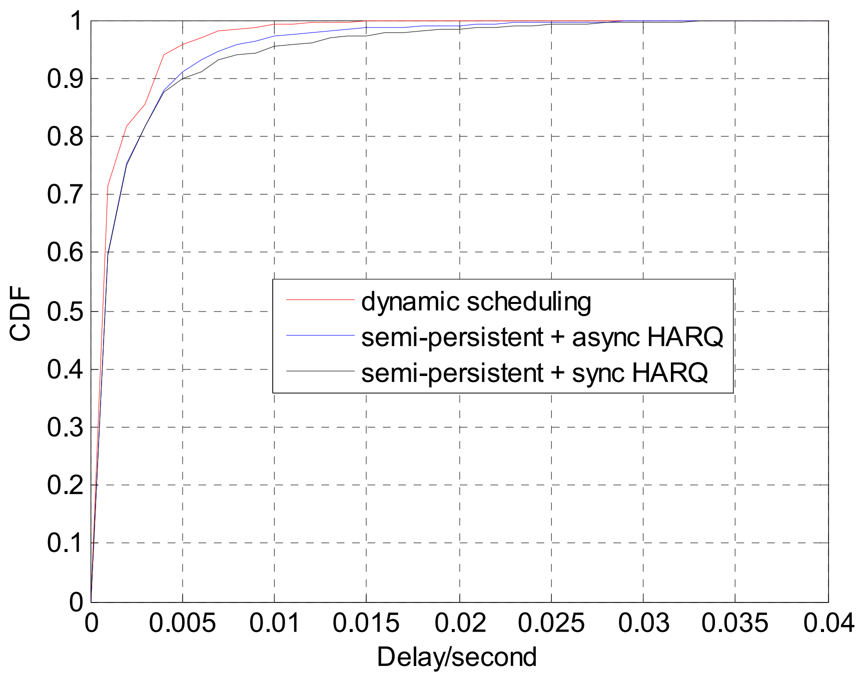

Figure A5 Delay CDF for 7.95Kbps AMR in Case 4 (with SID, 80 UE/sector)

A2.5 UL VoIP capacity for 12.2Kbps AMR in Case 4 (with SID)

In this case, SID packets are modeled and dynamically scheduled. From Figure A6 UL VoIP Capacity for 12.2Kbps AMR in Case 4 (with SID), 70-73 12.2Kbps users per sector can be supported. There is a capacity reduction of about 15% compared to 82-86 7.95Kbps users in Figure A3. The related delay distribution for different scheduling methods is shown in Figure A7 Delay CDF for 12.2Kbps AMR in Case 4 (with SID, 80 UE/sector), which is worse than delay distribution for 7.95Kbps AMR (Figure A5).

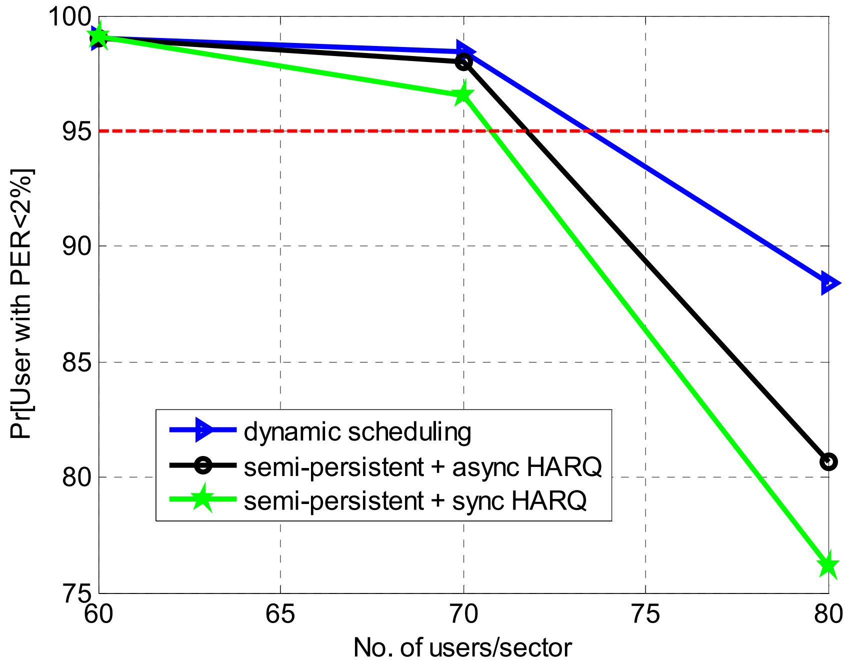

Figure A6 UL VoIP Capacity for 12.2Kbps AMR in Case 4 (with SID)

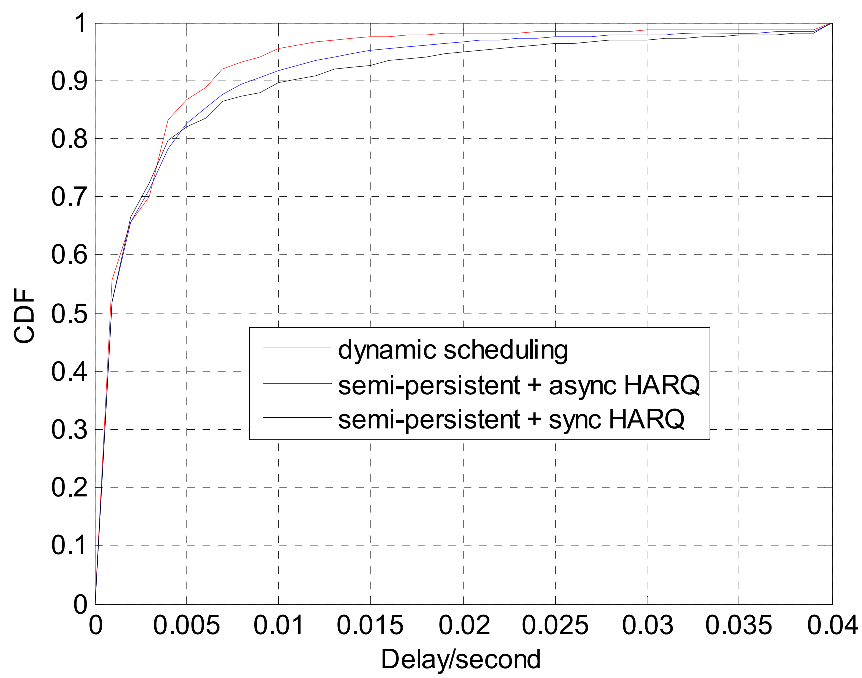

Figure A7 Delay CDF for 12.2Kbps AMR in Case 4 (with SID, 80 UE/sector)

A2.6 UL VoIP capacity for 7.95Kbps AMR in Case 1 (with SID)

In Case 1, for 7.95Kbps AMR and with SID modeling, about 78 users/sector can be supported (see Figure A8 UL VoIP Capacity for 7.95Kbps AMR in Case 1 (with SID)), so there is nearly 10% capacity reduction compared to 82-86 users in Case 4. The related delay distribution is shown in Figure A9 Delay CDF for 7.95Kbps AMR in Case 1 (with SID, 80 UE/sector), which is a little worse than delay distribution in Case 4 (Figure A5).

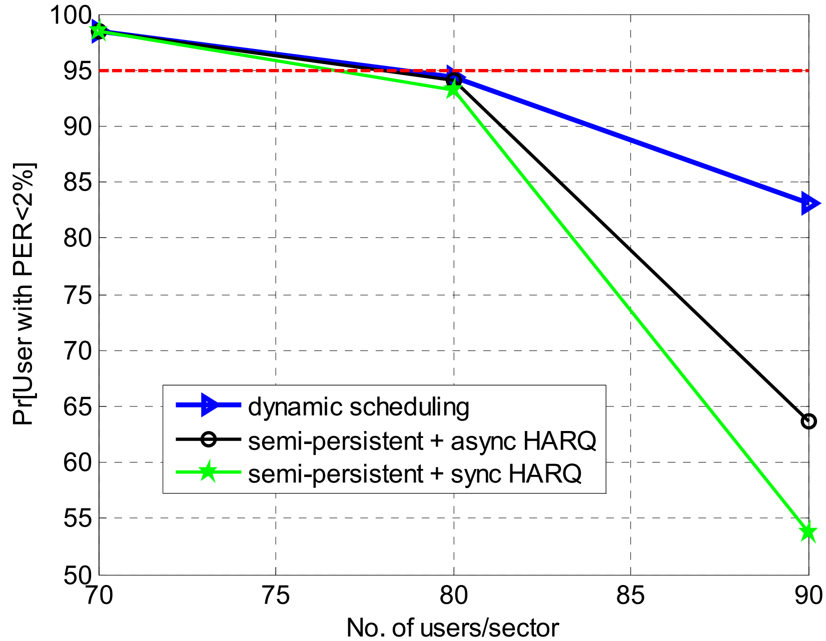

Figure A8 UL VoIP Capacity for 7.95Kbps AMR in Case 1 (with SID)

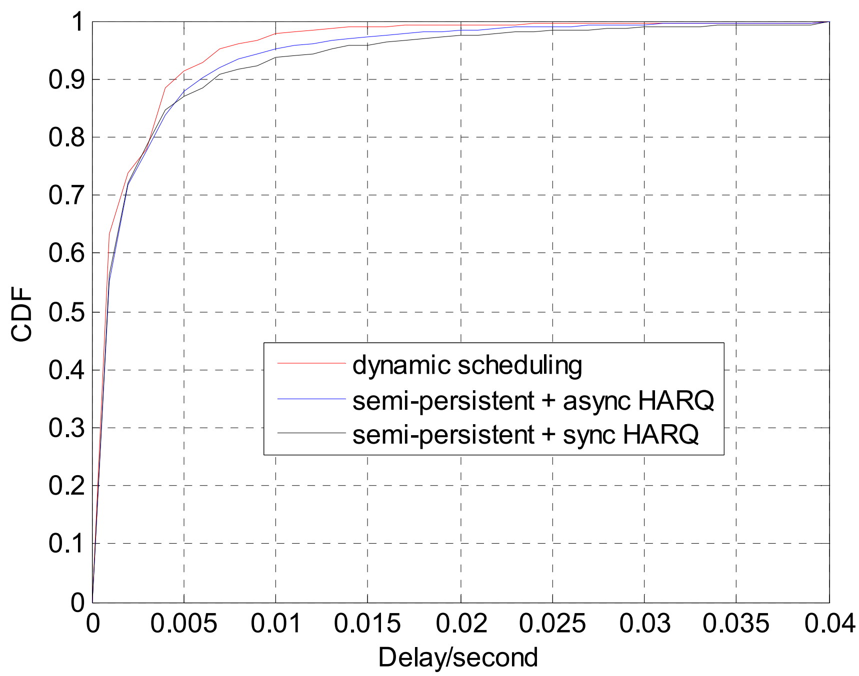

Figure A9 Delay CDF for 7.95Kbps AMR in Case 1 (with SID, 80 UE/sector)

A2.7 UL VoIP capacity for 12.2Kbps AMR in Case 1 (with SID)

In Case 1, for 12.2Kbps AMR and with SID modeling, about 63 users per sector can be supported (see Figure A10), so there is nearly 20% capacity reduction compared to 78 users for 7.95Kbps AMR in Figure A8.

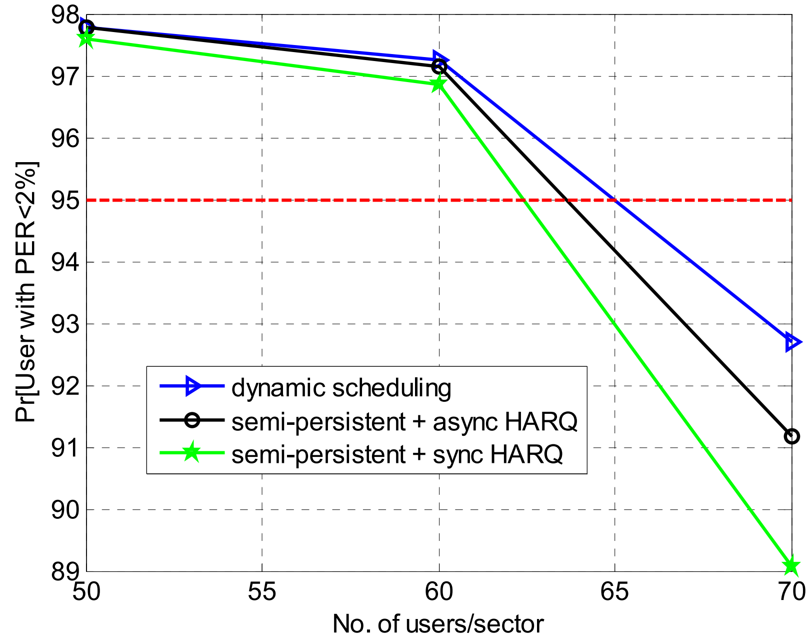

Figure A10 UL VoIP Capacity for 12.2Kbps AMR in Case 1 (with SID)

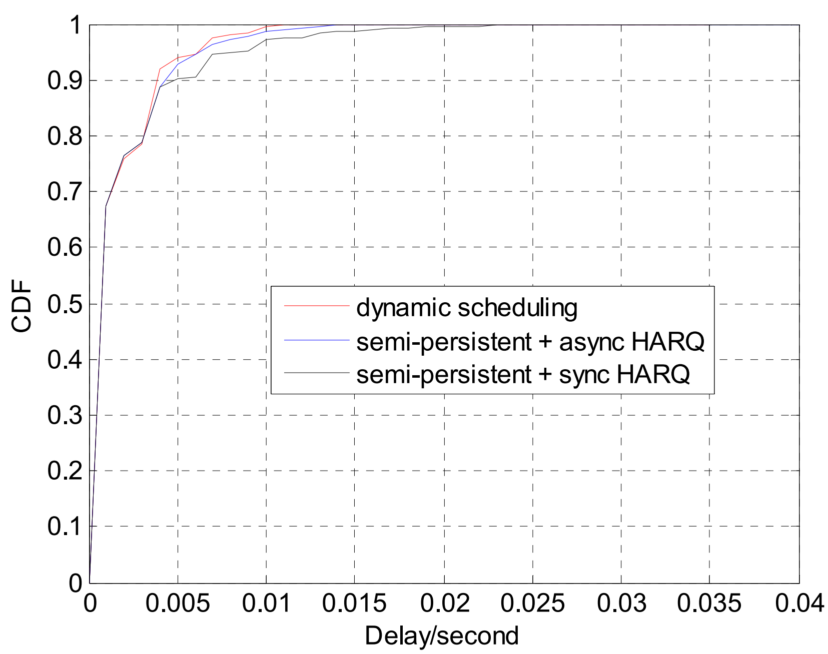

Figure A11 Delay CDF for 12.2Kbps AMR in Case 1 (with SID, 60 UE/sector)

A2.8 UL VoIP capacity for 12.2Kbps AMR in Case 1 (with SID, 5MHz BW)

Herein, we study the VoIP capacity with control channel reservation and restriction of number of scheduled users per TTI in Case 1 and 5MHz BW. There are totally 25 RUs in 5MHz BW and 4 RUs are reserved for control channel, so 21 RUs are left for data transmission. For dynamic scheduling, we simulate two cases: 1) we get the capacity without restriction of number of scheduled users per TTI. 2 RUs are for transmission of one packet, and then with 21 RUs for data, maximum 10 packets can be scheduled per TTI (1 RU wasted); 2) we get the capacity with restriction of maximum 7 scheduled users per TTI when considering of the DL allocation signaling overhead, and 3 RUs are for

transmission of one packet, and then with 21 RUs for data, just 7 packets can be scheduled per TTI. For semi-persistent scheduling, 2 RUs are for transmission of one packet and there's no restriction of number of scheduled users per TTI due to its greatly reduced signaling compared with dynamic scheduling.

The results are shown in Figure A12. With the restriction of number of scheduled users per TTI, dynamic scheduling can only support 196 users per sector; yet without any restriction of DL control signaling, up to 240 users can be supported for both dynamic scheduling and semi-persistent scheduling. However, no restriction of DL control signailing is unrealistic for dynamic scheduling. So semi-persistent scheduling seems to be very promising with high capacity and reasonable control overhead.

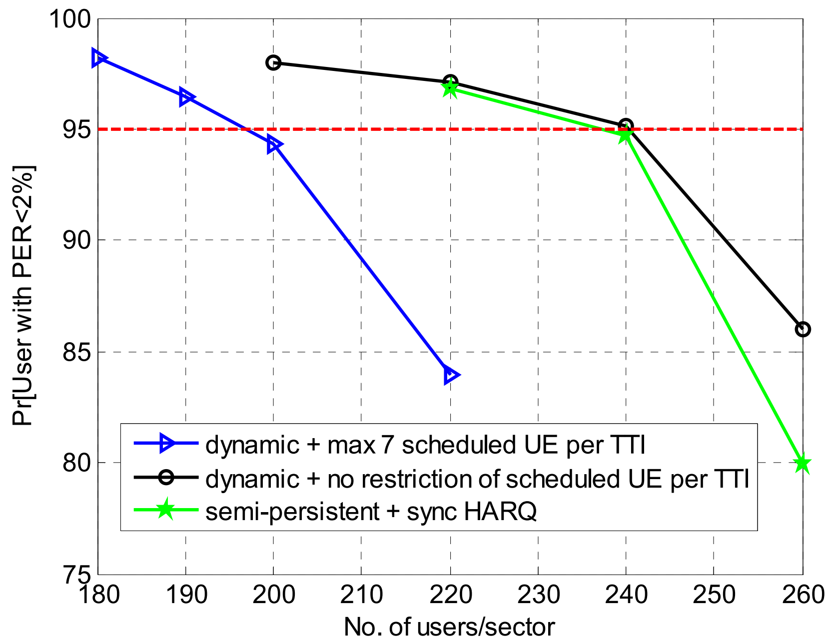

Figure A12 UL VoIP Capacity for 12.2Kbps AMR in Case 1 (with SID, 5MHz BW)